VOLUME 66
NUMBER 8

WHOLE NO. 340
1952

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Factor Analysis of Some Reasoning Tests

By

HAROLD M. CORTER

North Carolina State College of Agriculture and Engineering, Raleigh, N.C.

Based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Pennsylvania State College, 1949

Accepted for publication March 5, 1951

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1952, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

TABLE OF CONTENTS

I. *The Experiment* 1

II. *Description of Tests* 4

III. *Summary of the Findings* 17

IV. *A Discussion of Test Composition* 22

V. *Clinical Validation of Factors* 25

VI. *Summary and Conclusions* 27

Bibliography 30

ACKNOWLEDGEMENTS

GRATEFUL acknowledgement is made to the many persons who have played a role in aiding the completion of this dissertation.

Special thanks are due to Dr. Robert G. Bernreuter, Professor in the Department of Psychology, Pennsylvania State College, who served as chairman of my graduate committee. The other members of my committee, Dr. B. V. Moore, Head of the Department of Psychology; Dr. J. E. DeCamp, Professor in the Department of Psychology; Miss Mary Jane Wyland, Professor in the Department of Education; and Dr. W. U. Snyder, Associate Professor in the Department of Psychology, were helpful.

Dr. Nathan Jaspen contributed from his broad knowledge of statistics specific help in overcoming some of the hurdles involved in factor analysis. Fellow students helped by serving as guinea pigs for portions of the experiment.

Particular acknowledgement is made to the Junior class of the Bellefonte High School, Bellefonte, Pennsylvania, who served as the experimental group for this study. The staff members of this high school were very cooperative during the testing period and the author is greatly indebted to them. The two guidance counselors, namely Mr. Frank E. Hench and Miss Margaret Ferree, aided greatly.

H. M. C.

Raleigh, N. C.
August 8, 1950

Chapter I

THE EXPERIMENT

A. Introduction

The prime purpose of this study is to carry out an exploratory study in the area of reasoning, particularly in regard to various tests in the repertoire of the clinician. In terms of science in general, the procedure of factor analysis is one in which certain of the relational characteristics of reality are given some order and reference. In some ways it offers some hope that eventually psychologists will be able to speak in terms that other psychologists understand, and it may even some day result in a broad common ground through which all scientists can interchange ideas. Perhaps the most philosophically sophisticated of the people who have shown interest in factor analysis, Burt, defines the procedure in general as a means of providing parsimonious descriptions of psychological data (6, p. 249). Although Thurstone maintains that his search for "primary mental abilities" is not an attempt to resurrect an elementarism in psychology, in many ways he speaks about his factors as if they were such elements. In general the viewpoint is held by the present author that the mathematical analysis involved yields only data which have no inherent meaning and which must be interpreted by evidence from observation by trained people.

Two studies by Halstead (14) and Meili (22) present factors which, while somewhat abstruse, perhaps too highly abstracted, difficult to understand, and not completely identified, are ones that perhaps may eventually have more psychological usefulness than the highly objective test-material type factors now generally recognized. Highly objective factors in the traditional Thurstonian analysis have found some usefulness in the area of vocational guidance but have been somewhat sterile in terms of understanding either other individual behavior or its abnormalities. Descriptive terms which are satisfactory in terms of scientific knowledge should not only meet a criterion of objectivity in a parsimonious manner but should be able to contain or explain a broad range of human behavior if they are to be given the title of "primary mental abilities."

Prior to 1940 the results of factor studies have been subjected to a critical review by Wolfle (42). Since that time a well-organized summary of factor studies having relationships to aptitude and achievement has been undertaken by French (10). His analysis summarizes most of the studies made with young adults and adults. He makes an attempt to bring together various factors identified by different workers and renames them at times for cross identification. All in all, this book represents the contributions that factor analysts have made to our knowledge of psychology in digested form. The reader is referred there for further study.

B. Problems

One of the questions on which it was hoped that this study would throw some light is: Can reasoning be satisfactorily subdivided into separate factors?

Under this general question a number of possibilities might be true. Among these is the assumption that reasoning can be defined factorially in terms of the

steps in the reasoning process. These are generally given in terms of problem-solving ability in which the successive steps are: first, recognition of the problem; second, bringing ideas to bear upon the solution of the problem; third, the weighing of various solutions to the problem; and finally, the testing of such solutions in practice. For many years logicians have separated reasoning into two components: deduction and induction. It would be of interest to know whether or not these a priori systems of analysis are meaningful to the extent that they can be recognized in terms of the statistical procedures used in factor analysis.

The design of the experiment makes it possible to make some empirical evaluations between typically objective test-centered factors as *"verbal," "spatial," "numerical,"* etc. against more generalized factors as found by Meili. The test battery contains five of Thurstone's tests which are supposedly relatively pure measures of factors found in his work. The tests also contain in the portion of the battery which is individually administered a number of tests which are similar to those used by Meili. Among these are the Comprehension test (No. 7 in the battery), the Naming test (No. 11), the Block Design test (No. 17), the Picture Titles test (No. 15).

Another question was concerned with the factorial content of tests not previously included in factor studies, most of which were tests which purported to measure reasoning, as, for example, the Arthur Stencil Design Test (No. 17), the Holsopple Concept Formation Test (No. 19), the Hanfmann-Kasanin Concept Formation Test (No. 20), and the concept-formation tests devised by the author. Another general question to be explored by the experiment was one concerning the relationship of concept-formation tests to the reasoning process in general.

More than ten years ago Allport, according to Wolfle, questioned factor analysis because of its inability to meet a criterion of clinical usefulness or importance. Allport, in general, regarded factors at that time as statistical artifacts which have little or no relationship to real people. This complaint is still one which is felt to have some justification because of the rather limited usefulness of discovered factors. One aim of the present study is to bring the clinical and statistical approaches into somewhat closer alignment.

The largely rational, empirical approach used by clinicians perhaps needs some overhauling in terms of statistical strictness to prevent a meaningless verbosity; whereas, the highly objective statisticians need the application of empirical evidence from the clinic to make their manipulations have meaning. The viewpoint of the author is that the clinician needs to become a little more rigid in his use of language for the assessment of behavior, while the statistician needs to become more the psychologist and somewhat less the mathematician.

C. The Preliminary Investigation

A preliminary investigation utilizing a sample similar to the one to be used in the main investigation was undertaken in order to insure that the tests were of suitable difficulty, that they contained no errors or ambiguities, and that scoring methods and standards were suitable. A superficial examination of the results of this experiment was suggestive of fair possibilities in the group of tests involved.

D. The Population

The junior class of a high school in a town of approximately 7,000 population formed the basis for the selection of the experimental population. The Thurstone Primary Mental Abilities Test, Science Research Associates Edition, was administered to the entire junior class during one day of testing in the school. From the students present on the day of testing, a random selection was made and a list of 110 names compiled. From this group the 100 cases used in the experiment were selected in order. During the course of the experiment two students left school, one was not present on the final day of testing, so 103 names were used. Junior-class membership was divided into 46 per cent girls and 54 per cent boys. In the 100 cases used there were 57 boys and 43 girls.

This high school draws its population about equally from the town and townships that are adjacent to it. In the experimental group 43 pupils are transported by bus to the school. This is a rough way of estimating the number who are from rural areas.[1] The mean age as given by the pupils on the test blanks, where they were requested to record their ages to the nearest half year, was 17.46 years. This may seem slightly high for high-school juniors, until it is recalled that testing was conducted during the last two months of the school year. From the standpoint of achievement and maturity the group occupied a position that was midway between a junior- and senior-class group.

E. Test Administration

Following the scoring of the paper and pencil tests and the selection of the sample, the high-school guidance advisors were requested to select four students, two boys and two girls, among the first fifty on the list whom they regarded as leaders in the class. These four persons selected were the first four to whom the individually administered tests were given. The purpose of this selection was to secure the cooperation of the remaining students to be later tested. Following the testing of these four persons, the remainder of the subjects was taken in convenient order. The place of testing was satisfactory, although upon occasion testing would be interrupted when movies, band practice, and other disturbing programs were held in the auditorium. All the testing, except for the final 20 subjects, was conducted by the author. With the final 20 subjects tested, the first 8 of the 16 individual tests were administered by a trained worker. With the first five subjects not tested by the author close supervision of testing procedures was made to insure standard administration.

In general, the test battery appeared to be appealing to the students as evidenced by the generally excellent rapport. How much of this was motivated by materials used in the test and how much was the result of students wanting to get out of regular classes was a question, but the over-all result was quite satisfactory. The average time of administration of the entire individual battery was approximately two hours at the beginning of the experiment. As the test material became more familiar, the examiner's efficiency rose to a point where, in most cases, the testing was completed in an hour and forty minutes, or two regular periods of the school session. In general, this period of time was

[1] These data are based on attendance records on file in the school.

sufficient to allow a pace that was unhurried but not wasteful of time.

F. The Individual Test Battery

In general, it was felt that the individual tests should meet certain criteria in order to qualify as suitable tests. Owing to the fact that the examiner had to conduct the testing at rather irregular intervals, it was necessary to include only tests that were easily transportable. Ease of administration, objective scoring, and satisfactory scoring standards were other criteria used. Still another qualification of the test battery in general and of the individual tests themselves was that they should be administerable in a reasonably short length of time. In most cases, the latter was achieved to the point where no individual test in the battery consumed more than ten minutes. Tests which did not meet this criteria in their original form were shortened or modified. Another criterion was suitability in terms of difficulty. These criteria were, by and large, established in the preliminary study with a comparable group.

In order to insure comparable measures for each student and to minimize artifactual gains in performance due to gains in maturity, the testing altogether covered a period of only fifty-four days.

Chapter II

DESCRIPTION OF TESTS

The final test battery consisted of 21 tests, 5 of which were the subtests of the Science Research Associates version of the Primary Mental Abilities Test of Thurstone for group administration (Intermediate – Form AH [38]). The remaining 16 tests were all of an individual nature. They consisted, in general, of word tests, perceptual tests, and tests using various concrete objects. A more thorough description of the entire battery follows:

Test No. 1. SRA PMA Verbal test.
Test No. 2. SRA PMA Space test.
Test No. 3. SRA PMA Reasoning test.
Test No. 4. SRA PMA Number test.
Test No. 5. SRA PMA Word Fluency test.

These tests were included in the battery, largely for purposes of factor definition and to provide a means of comparing some traditional objective "Thurstonian" factors with Meili's factors.

An examination of the succeeding tests in the battery will reveal that all these tests except No. 4 had counterparts in the individually administered group. Test 1 and Test 8 were both vocabulary tests, and Test 10 had vocabulary elements. Test 2 had tasks felt to involve similar elements in No. 13, No. 17, No. 18, No. 19, and No. 20. Test 3 had, of course, most tests resembling it in the reasoning area. Test 5 was felt to be quite similar to No. 11, with some resemblances to No. 15.

Test No. 6. The Wechsler Form II Similarities test (40).

This test was included in the test battery to give a measure of verbal concept formation. Another reason was that it was felt that it would be valuable in establishing rapport, and readily accepted by the group. This proved to be the case.

Test No. 7. Wechsler Comprehension Form II (40).

This test was included in the battery because of its similarity with a series of practical questions mentioned in the Meili study and because it, too, was felt to be quite acceptable to most subjects and a good introductory test.

Test No. 8. A modification of the Stanford-Binet Vocabulary Test (27). The items from No. 14 to 38 inclusive of this test were presented. The instructions for the test were modified slightly and were as follows:

"I have here a series of hard words. Look at each word and tell me what it means. What does 'skill' mean?" Otherwise, administration followed usual Stanford-Binet standards.

This test was included in the battery to give added strength to any *"verbal"* factor and to compare the two different methods of test administration, group and individual, in terms of the factors in vocabulary tests. It was felt that the individually administered vocabulary test

would require a combination of fluency and word knowledge.

Test No. 9. Judgment test. This test consisted of a series of ten cards, each with a question and four suggested answers, which the subjects were required to rank in order of which answer appeared to be best.

A preliminary selection of 20 items was made from Cardall's Test of Practical Judgment (7) that was felt to be suitable for students of a high-school level. From this group of 20 items, the ten which had the greatest weights according to Cardall's scoring standards were finally selected. These ten items were Nos. 10, 12, 15, 19, 20, 26, 37, 41, 44, and 47.

The instructions for this test were, "I have a series of cards that have questions on them. This test is to see whether or not you have good judgment. Look at each card and tell me which answer you think is best, which is second best, and which is third best. Try this one." After the first card the instruction was given, "You need only to tell me the correct letters in the correct order." (This was said if the subject gave the complete phrase rather than the letter.)

Scoring used Cardall's weights for these items.

This test was used in the battery because of its unknown factor composition and because some judgment measure that had face validity was felt necessary for definition and evaluation purposes.

It was felt, too, that it would provide a rather differentiating measure, and would serve to demonstrate the validity of the discrimination element in the reasoning process.

Test No. 10. Word Sorting test. This test was devised by the author to provide a measure of verbal concept formation. The test consists of a series of 12 cards, on each of which there are eight words, which can be divided into two groups according to a principle. In the first eight cards, the set which the subject has must alternately change from meaning of the words to some inherent characteristic such as the letters in the words, etc. A preliminary study was made with this test in an attempt to eliminate the possibility of having more than one correct response or method of sorting for each set of words. This attempt was successful, for no other means of classification were given by the experimental group which met the sorting requirements.

The directions for the word sorting test were as follows:

"I have here a series of cards with 8 words on each. On each card there are always two different kinds, or groups, or classes of words. What I want you to do is to tell me what two kinds there are on each card. Here is the first one. (Present demonstration card.) It will help you if you remember there are always four words of each kind on each card. It may be anything about the words; what the words mean; the way the words are written, or anything at all." At card No. 8 the instruction was repeated, "Remember it may be anything." The first card was explained if the subject failed to give the correct response or if he apparently had misunderstood the instructions. Similarly, when the first error was made, the criteria by which the words had to be sorted were reviewed for the subject. Most of these errors consisted of classifications with more or less than exactly four words in each group. Another question that was frequently necessary to ask was, "What is the other kind?" This occurred when the subject only classified one of the two groups on each card. The list of words was as follows:

1. FAT, hat, mat, pat, RAT, BAT, cat, SAT (Demonstration)
2. Betty, Paul, Ralph, Charles, Harriet, Helen, Harold, Florence
3. hay, height, obey, bite, sleigh, weigh, kite, right
4. beef, cow, hog, veal, deer, mutton, venison, sheep
5. race, runt, sack, sick, rope, rule, sole, sunk
6. John, Henry, Man, Woman, Person, Child, Genevieve, Susan
7. so, what, goes, no, or, duck, to, fill
8. stinking, sweet, foul, slimy, pretty, nasty, melodious, pleasant
9. wierd, slowly, silly, wiegh, funny, realy, Febuary, seldom
10. doves, children, men, oxen, muskrat, elephant, snake, eel
11. sad, lad, cad, fad, bad, mad, glad, dad
12. egg, look, bed, up, odd, silly, light, rang
13. editor, moon, lax, acting, betting, uncle, odd, many

In the scoring of these items, one credit was given for each group correctly categorized. Since each card had two groups of words alike in a certain respect, a total of 24 credits was possible. The subject was permitted to try the card as long as he appeared to be productive. Generally after a minute or so he would say, "I can't see anything there." In such case, he was instructed to try the next one. Testing was concluded after four successive item failures. In some cases a subject was able to name four of the words which belonged to one group and to give the principle by which they were grouped, but was unable to give the principle for the other four. In such cases, a credit of one was given.

On the sample item the two groups of words are words written in capital letters and words written in small letters. In the second item, four of the names are women's names and four are the names of men. The similarity in sound provides the basis for sorting of the third item, since there are four words which have an *ite* ending and four words which have an *ay* ending. In item four, four of the words are names of animals

and the other four are the names of kinds of meat. In item five, four of the words begin with *r* and four begin with *s*. Proper nouns and common nouns or equivalent expressions thereof are the two groups used in item six. Four-letter words and two-letter words are the two groups in item seven. In item eight, four of the words have pleasant connotations while the other four are unpleasant. (This was mostly expressed by the subjects in terms of "nice" and "not nice.") In item nine, four of the words in the group are purposely misspelled; the other four are spelled correctly. In item ten, four of the words are singular and four are plural. Adjectives, or describing words, and nouns, or names of things, are the two groups in item eleven. Four of the words in item twelve contain double letters, while the other four have no letters repeated. In item thirteen, four of the words are begun with consonants and the other four with vowels.

This test was included in the battery to give a sorting test using verbal material that would be analogous to the Buttons test, the Opposites test, and to the various block-sorting tests used. It was felt to combine elements of verbal factors and the hypothesized reasoning factor or factors.

Test No. 11. Naming test. This test was another devised by the author in an attempt to measure a crude sort of concept formation with certain deductive elements, again using verbal material. It has been noted that early concept formation is apparently built around the abstractions made in the use of language, the assigning of material to various classes.

The instructions for the test were, "In this test, I am going to ask you to name all the things of a certain kind that you can think of in one minute. For example, I might ask you to name *white* things, and you would say 'snow,' 'clouds,' 'linen,' or 'paper.' Try this one. 1. Name as many red things as you can." The test was then continued with 2. round things, 3. sharp things, 4. cold things, 5. soft things, 6. elastic things, 7. things that there are generally five of in a group. During the testing, at the first pause in each group, the subjects were urged to continue by saying, "anything that is——will do."

One credit was given for things that are always or usually associated with the named characteristic. One-half credit was given for things that are occasionally or sometimes associated with the characteristic. The total score on this test was the total number of credits in all the categories mentioned. Examples of one-credit scores in the first category are fire-trucks, stop-lights, and finger-nail polish; half-credit items were such things as cars, houses, dresses, etc. In the second group, all things that have the general characteristic of roundness were given full credit, including items in which the shape was primarily spherical. The things given one credit were such things as rings, circles, etc. Half-credit answers here included such things as compacts, tables, etc. In the third group, both things that have sharp edges as axes, saws, etc. and things which were pointed as pencils and ice picks were given credit. One-credit items are like those mentioned above, whereas half credit was given for such things as pieces of metal, etc. Cold things were credited in which the association is made of usually being cold as ice cream, snow, ice, etc., and half credit was given for such things as water, ground, or air. Soft things in the full-credit group included such things as pillows, sofas, mattresses, etc.; and in the half-credit group such things as food, chairs, etc. Elastic things included such things as rubber bands and balloons in the full-credit group, and such things as socks and other articles of clothing in the half-credit group. In things that there are generally five of in a group, such things as cents in a nickel, or men in a basketball team, or tires on a car (including the spare) were full-credit items. Such items as petals on flowers, or leaves in a group were given half credit, if the specific flower or leaf having these characteristics was not named.

This test was included in the battery to provide a measure of fluency, wealth of ideas, and deduction.

Test No. 12. Opposites (see Illustration 1). This test was devised by the author as an attempt to get at certain aspects of concept formation in a task which involves looking at drawings of various objects. The final group of drawings selected included four test cards plus one demonstration card. The cards were called circles, trees, squares, letters, and stamps. The task was to select from a series of five choices an object which was most unlike another object that was first presented on the card. In order to accomplish this, the first step in the process was to attempt to define the way in which the objects varied, and then to select the object which varied in the greatest number of ways. On the demonstration card two variables were given. A small black circle was followed by three numbered circles which were successively small and white, large and black, and large and white. Another item for additional practice was given on the demonstration card. On each of the four successive cards, which made up the actual test, there were four items which varied in terms of three variables per card, and the selection was made from five numbered items. On the trees card, the variables were the shape of trees (either rounded or triangular shaped), tall or short, and having light or black trunks. On the second card, squares, the variables were color (either red or white in the outer portion of the square), size of insert, and shape of insert (either a diamond or square in a rotated or unrotated position).

On the letters card, the variables were whether or not the letter was an *a* or a *b*, whether or not

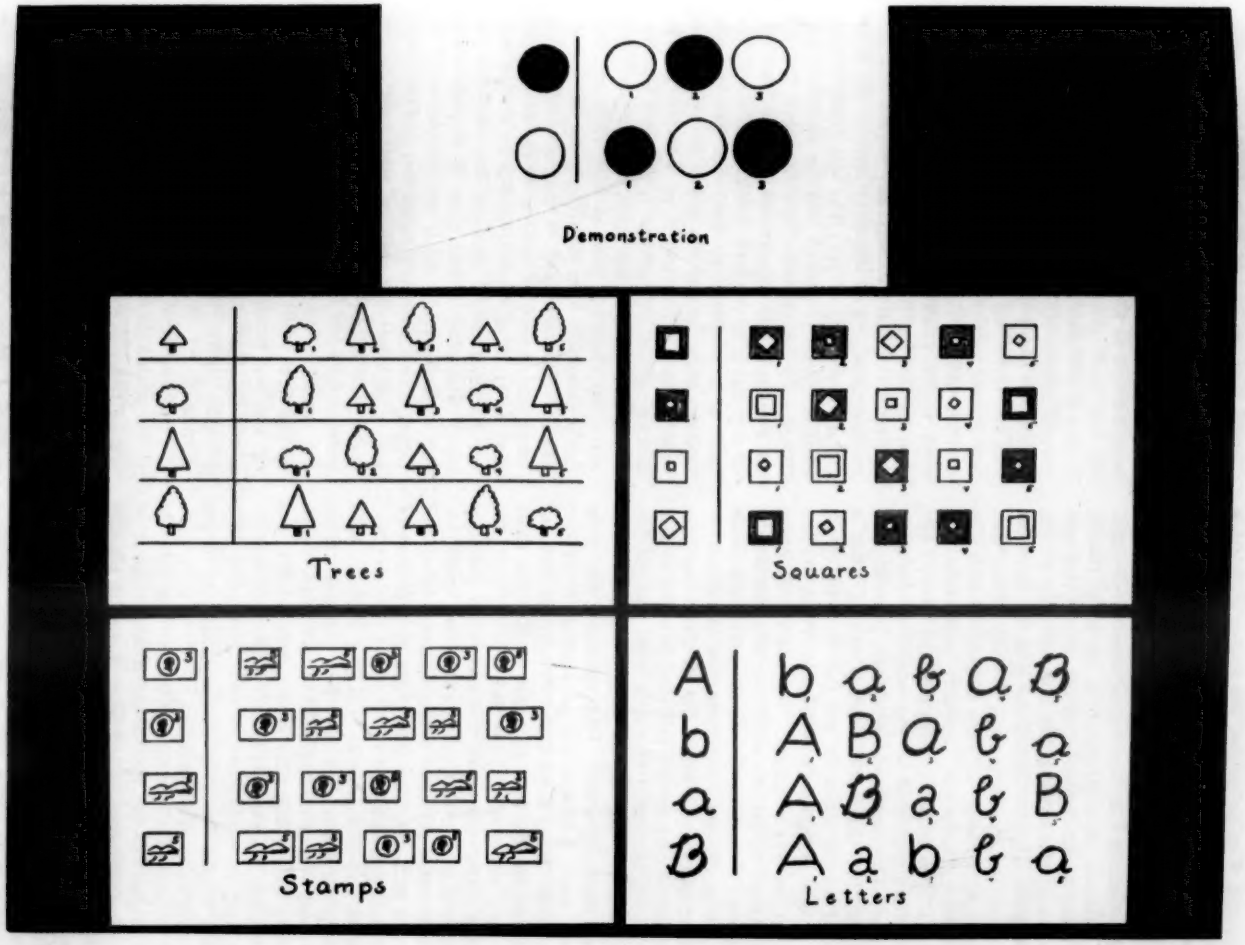

ILLUSTRATION 1. Opposites Test.

it was printed or written, and whether or not it was capitalized or a small letter.

On stamps, the variables consisted of length (long or short), the number on the stamp (5 or 3), and whether or not the stamp depicted a scene or a person.

Instructions were as follows: "Here is a series of cards that have various things on them (the first card was presented). Notice the first circle here is followed by three numbered circles that are different from it. Point to the circle that is *most unlike* the first one." If the correct response was given, the subject was reassured by saying, "Yes, that is right." Or, if he gave an incorrect response, the examiner said, "No, it is this one, because this one is not only larger, but it is also black," or made the necessary correction and added, "It is unlike in both size and color." The other item was tried by the subject in the same manner. The next card was then presented by saying, "Try this one." During testing, when a subject would begin to tell the ways in which the selected response was different from the test subject, he was informed, "Tell me the numbers of the ones on the cards that are most unlike the first one." Following the conclusion of this phase of testing, the cards were returned to the subject and he was asked, "What are all the things that you had to look for on this card in order to find the right answer?" The first time one of the variables was omitted, the examiner asked, "Was there anything else?" This question was only asked once. Occasionally a repetition of the question, "What are all the things you have to look for on this card in order to find the right answer," was made.

One credit was given for each item correctly answered in the first group, plus one credit for each variable correctly named during the second part of this test. The total score was the sum of the two types of scores.

This test was included in the battery as a concept-formation test that had the anomalous characteristic of finding differences in addition to finding similarities. Its factor composition was unknown, but was felt to revolve around reasoning and it was included as a perceptual task which would measure concept ability by means of recognition of difference.

Test No. 13. Gottschaldt Figures. This test was an adaptation for individual administration of the first 12 items on the SRA Research Edition of this test for group administration (31). These items were cut out of the Research Edition and pasted on 3 x 5 cards with the preliminary figure located to the left of the four succeeding figures in which it is to be found. For demonstration items the instruction page figures were used. These instructions were presented on three different cards which became successively more difficult. The items were then numbered for convenience in tabulating. Thurstone has described the characteristic that this measures as the ability to break perceptual gestalts (29).

The instructions for the test were as follows: "I have here a series of cards on which some designs are hidden in some more complicated designs. Look at this card (the first card in the demonstration series is presented). As you see, all of the designs here in this row have a square in them (pointing to the squares). I have made pencil marks so that they stand out more clearly. In this next group, designs No. 2 and No. 3 have a *V* like this in them (again pointing to the correct responses). This one has a *V* too, but it is not the same shape (pointing to No. 1), and this one (pointing to No. 4) has a *V* too, but it is not turned in the same direction, so that is not correct, either." (Card No. 2 is presented.) "This card is for practice. Which numbered designs contain the big plus sign like this?" (After the subject's response, correct any errors and give explanations.) "Try the next one on the card." (Again correcting errors.) The last demonstration card was then presented and the examiner said, "Here is a harder one like the ones you will have to do in the regular test. Try this one. Part of this test is to see how quickly you can do this kind of a problem, so you will have to work rapidly. Don't spend too much time on any of them. If you can't see it in number 1 rather quickly, go on to number 2, etc. You will have thirty seconds for each card, and this will generally be time for you to look each one over fairly carefully. You need only to tell me the numbers for each card. It may also help you to know that it may be in none of the designs, or it may be in all of them. There is no certain number on each card." (Present last demonstration card.) The subject then gave his response to the last demonstration card, any errors were corrected, any omissions pointed out and the subject was cautioned, "Sometimes you have to look pretty carefully." The first test card was then presented and at the conclusion of thirty seconds, the examiner suggested that the subject go on to the next card. On succeeding cards, however, he was not prompted to try the next card and generally chose his own pace, except in cases of extreme delay, where the time exceeded one minute.

The number of items correctly given according to the standards set forth in the examiner's manual for this test was used.

No time scores and no limitations of time were made except for the instructions. These instructions were designed not to provide a "speed" test but to prevent a subject dallying too long over each card. It was felt that, given sufficient time, all subjects could answer correctly, otherwise.

This test was included in the test battery because it was felt that it might have some connection with reasoning in the analytical ability required to perform satisfactorily and give some

idea of the subject's ability to break a particular set.

Test No. 14. The McGill Picture Anomaly Series (17), Form N, the abridged version. This test was a series of picture anomalies or absurdities consisting of 21 items in the abridged form N.

Instructions for the test were simply, "Point to what is funny or out of place in these pictures." Additional instructions, "Use this brush so you don't make any marks on the paper," and "We'll only do the ones marked with a star or asterisk beside the number," were included.

This test was included in the battery because it was felt that it might play some role in delineating the role of rejection of hypotheses in the reasoning process, in which a large number of hypotheses are brought to bear upon the problem and the absurd ones are discarded. The recognition of error in general was felt to be of some importance in reasoning and its role needed definition.

Test No. 15. Picture Titles. This test suggested by one of the items in Meili's battery (22) consisted of a series of five pictures to which the subject was requested to give titles. There was also one demonstration picture. The pictures were from Wells and Ruesch's *Mental Examiner's Handbook* (41). They are *Bus Passengers,* by Isaac Soyer; *Morning on the Cape,* by Leon Kroll; *After the Tornado, Bahamas,* by Winslow Homer; *The Pioneer,* by Winslow Homer; *Rum Cay,* by Winslow Homer; and *The Nesters,* by Tom Lee. The *Bus Passengers* picture, by Soyer, was used as a demonstration card.

The instructions for the test were as follows: "I have here a series of pictures to which I want you to give titles. Give three titles for each one. (Present the first card.) For example, you might call this card, 'People,' 'At the End of the Day,' or perhaps 'Bus Riders.' Try this one." If he failed to give three responses, he was urged by the words, "Give three names for each picture."

Two credits were given for a title which subsumed the basic characteristics of the picture. One-point credit was given for titles which consisted of mere description or enumeration. On the first card, satisfactory two-credit responses were such things as "Morning on the Farm," "Beginning of the Farmer's Day," etc.; one-credit responses were given for such things as "School Girl," "Ploughing," etc. On the next card, "After the Storm" and "Shipwrecked" were two-credit answers, while "Desert Island" and "Dead Man" were examples of one-credit answers. On the next card, "Clearing the Forest" and "Pioneer" were given two credits, while such titles as "Woodchopper" and "Lone Tree" were given one credit. On the next card, two-credit responses were such things as "Catching the Turtle" or "Race for the Sea." One-credit answers were such things as "The Turtle" or "The Black Man." On the last card titles as "Western Farm Scene," "Pioneer Farm," and "Farm Out West" were two-credit answers, while "Farm Girl" and "Ploughing" were one-credit answers.

This test was included to give a measure of perceptual fluency or richness of ideas, or ability to synthesize the elements of the situation by abstracting the essential meaning. It was also felt, because of the request to give three titles, to give a measure of some of the power elements of intelligence.

Test No. 16. Rorschach (Beck's *Z* score) (4, pp. 58-82). In this test three of the series of Rorschach's ink blots, cards number II, III, and IX were presented.

The instructions for this test were "I have here a series of three ink blots. What I want you to do is to tell me what they look like to you, what they might be, and what they remind you of. Tell me three things for each card." The cards were then presented with the additional instruction that "You may use any part of the card, or turn the card any way you like to give your answers."

Subjects were urged to produce three responses by saying, "Look at it a little longer, perhaps you may see something else," when they stopped short of the requested three responses. Following the presentation of the three cards, an inquiry for position and further definition of the organizational elements required by the *Z* score was made. Other Rorschach inquiry was omitted.

Scoring was according to the standards established by Beck. The three card scores were then combined to get a total *Z* score.

This test was included in the battery as a somewhat less concrete test of the abilities required in the preceding test. Too, it was felt that the ability to make combinations was of some importance in the reasoning process, and according to Beck has definite intellectual aspects. If this were so, it was felt that they would revolve around some sort of reasoning ability. It was felt that it might give some test of Meili's belief that factors could be applied to all of behavior and not just cognitive functions.

Test No. 17. Arthur Stencil Design Test (2). In this particular test not all of the cards were presented because of the time limitation. (The total testing time for a complete test is usually about 30 minutes.) The final series of cards selected consisted of numbers 1, 4, 5, 8, 13, 17, and the demonstration card.

Another change was from a time limit of four minutes, which Arthur uses, to one of two minutes in order to give a test of suitable difficulty for high-school students. The instructions for this test were as Arthur suggests.

Three credits were given for each design completed within 2 minutes. Additional time credits were given as follows:

1.	5^s-1	
4.	20^s-1	10^s-2
5.	20^s-1	10^s-2
8.	40^s-1	20^s-2
13.	40^s-1	20^s-2
17.	50^s-1	35^s-2

These were then summed for a total score.

Only the necessary stencils to complete the required designs were presented.

This test was included in the battery to give some idea of its factor composition since it purports to measure, according to the author, "logical reasoning." Another reason was that it appears to be a task rather similar to the Block Design test No. 18 and its factor composition could be compared with that of a test that has been previously subjected to factor analysis. Another reason for its inclusion in the battery was the fact that it was felt to be a good transitional test from the perceptual tests previously used and the tests involving concrete objects used subsequently.

Test No. 18. The Block Design test. This test is the Wechsler Form II Block Design Test (40).

The inclusion of this test in the battery was suggested by the variety of factor composition found by different investigators using a variety of this test.

Test No. 19. Holsopple Concept Formation Test (21). This test consists of 16 different blocks which vary in eight ways, each block containing four of the variables. The variables used are color (blue or red), shade (light or dark) or intensity, shape (square or octagonal), width, height, size of insert, color of insert (white or yellow), and shape of insert (triangular or circular). The blocks can be sorted into four groups in a total of some 28 different ways, using two principles. This test is an unpublished test still in an experimental stage. While it is presented here as No. 19, it was actually preceded by test No. 20, which served the purpose of acting as partial instruction for this test.

The procedures followed in the instructions were as follows:

"I have here some different kinds of blocks. These blocks like the others can be sorted into four piles using two different principles. However, with these blocks there are always exactly four in each group, and they can be sorted in a number of different ways. What I want you to do is to sort them into four groups any way that occurs to you." (If necessary, additional instructions were given.) Following the first spontaneous sort, the subject was asked to verbalize the basis of the sort. If he only mentioned one variable, or had some difficulty in verbalizing any variable, further questioning was conducted by asking, "What is the way in which all the blocks in this group are alike?" If the subject still failed to recognize the sorting principle, he was asked, "How does this group of blocks differ from this group?" Following this period, the subject was requested to "Now sort them in a different way from the way you sorted them the first time." Inquiry was conducted as with the first sort. Then the subject was prompted to sort them by saying, "As you see, these blocks can be sorted in many different ways. I am thinking of a way that is different from those you've already used. Go ahead, and start sorting the blocks to see if you can figure out the way that I have in mind. You might make some mistakes at first, but after awhile you will see the way I have in mind and be able to do them without error." The last sort was made according to an a priori judgment by the author concerning the difficulty of various sorting principles. These were in order (including the principles of height and width used in the test previously administered, No. 20):

1. height
2. width
3. color
4. shape of block
5. shade
6. shape of insert
7. color of insert
8. size of insert

While Nos. 1 and 2 are ordinarily among the more difficult of the sorting concepts, it was felt that since they were the principles in the previous test, they would be among the easiest sorting principles on this test. If the subject used these principles, he was given full credit. Other principles were used with less frequency. During the first two sorts it was possible to utilize only three of the principles. In some cases, however, four were used and in such case, as in the case where the three were used, the next two levels of difficulty were presented in the final sorting. In cases where the sorts used were among the more difficult ones, the easier ones which had not been utilized were employed for the final sorts, except that only one variable out of the first two, height and width, was ever employed.

Four credits were given for each of the first two sorts correctly done without error. Two credits were given for sorting combinations which employed two principles in two of the sorting spaces and other principles in the other two spaces. In such cases, the subject was reminded, "Look them all over to see if you have sorted the same way in each group." If he spontaneously rearranged the blocks correctly, a score of three was given. On the third sort, a score of four was given if the subject recognized the principle of sorting one block beyond the point where sorting had become determinate. Three credits were given two blocks beyond. Two credits were given beyond this point. If the principle was discovered (based upon the criterion of errorless placement in the remainder of the blocks) before the last four blocks, one credit

was given. Sortings are possible in which the principle does not become determinate before the ninth block, and in such cases the one score was reserved for the last three blocks.

Scoring for the inquiry portion of the test was as follows: Four credits were given for adequate verbalization of both sorting principles. If it was necessary to ask the way in which all the blocks in the group were alike or to inquire as to the difference between groups, only one credit was given for each variable provided the subject was able to verbalize it adequately. No credit was given in cases where verbalizations were inadequate. The total score on the test was the sum of the sorting score and the verbalization score.

This test was included in the battery because of its unknown factor composition, and the fact that it purports to measure concept-formation ability. It is different from the next test (No. 20) in that it covers a much wider range of material and has a characteristic of flexibility lacking in the other. The testing procedure was very flexible and permitted a spontaneity not found elsewhere in the battery.

Test No. 20. Hanfmann-Kasanin Concept Formation Test (16). This test was modified by the author to save time and to provide a more objective scoring method. To make it a more practicable test, it was decided that the following directions and instructions would be given.

"I have here a group of blocks. There are four different kinds of blocks in the group. This one (picking up the yellow semi-circular block) is called a Bik. It goes here" (placing the block on the appropriate corner of a card divided into four spaces which was used as a sorting area). "It goes here where it says 'Bik.'" "This kind of block (picking up the large blue square block) is called a 'Lag.' It goes here" (placing the block in the Lag space). (Picking up the small white hexagon), "This kind is called 'Cev.' It goes here" (placing the block in the Cev space). "This kind of block (picking up the high narrow orange colored triangular shaped block) is called 'Mur'; it goes here" (placing it in the appropriate space). The examiner then continued, "Now I've shown you one of each of the four different kinds of blocks in the group. What I want you to do is to go ahead and put the rest of the blocks where you think they belong. Pick them up one at a time, be careful not to look at the bottom, and put them where you think they might go or belong. At first you may make some mistakes because you won't know the way the blocks are sorted. If you make a mistake, I will tell you where the block should go, and after a while you will be able to see the way the blocks are sorted and be able to do the rest without making a mistake. Go ahead now and try one where you think it might belong." Corrections were made as necessary, and the blocks were turned over to reassure the subject that the examiner was not making up a sorting scheme of his own at each mistake. This procedure was continued until all the blocks were sorted. The subject was then asked, "How do they go?" And if he hesitated, "What is the principle by which the blocks are sorted?" If he still was unable to see the sorting principle, he was finally asked how all the blocks in each particular group were alike. And if he still failed to perform satisfactorily, he was requested to point out the difference between groups by the instructions, "In what way are the Murs different from the Cevs?" If he still failed, the solution was explained to him. The latter phase of the testing was conducted largely as an introduction to the next test, No. 19.

The scoring scheme used in the test was a combination error-time score. The number of errors made was subtracted from 20 and the number of minutes to the nearest quarter was subtracted from five and multiplied by four. These two scores were then added for a total score.

This test was included in the battery because of its unknown factor composition, its fairly wide use in the clinical field as a test of concept formation, and the rather rigid requirements of its unique solution for its performance. In contrast with test No. 19, time elements were utilized in its scoring and the freedom of choice in the above test was not only unavailable here for the subject but was probably handicapping.

Test No. 21. Buttons (See Illustration 2). This test was devised by the author specifically to give a test that was similar to 10 and 12, using concrete objects rather than words or perceptual material as the test items. The test consisted of a series of ten cards (post card size) upon each of which there were a number of buttons which were all alike, except for one, in a certain way. An additional card used for demonstration contained six buttons which were identical except one. On the remaining cards the buttons were all different and the subject was required by the nature of the test to recognize the way in which all the buttons were alike and/or select the one that was different in a respect in which all the others were alike. On card No. 1 all buttons had the same diameter except one. On card No. 2 the buttons all had two holes in the center except one. On this button the holes were near the edge. On card No. 3 all buttons had two holes except one which had four holes. On card No. 4 all buttons were opaque or "colored" while the different button was transparent or colorless. On card No. 5 the buttons were all upside down and all possessed the common characteristic of having a projection for insertion of the thread that lay parallel to the face of the button except one

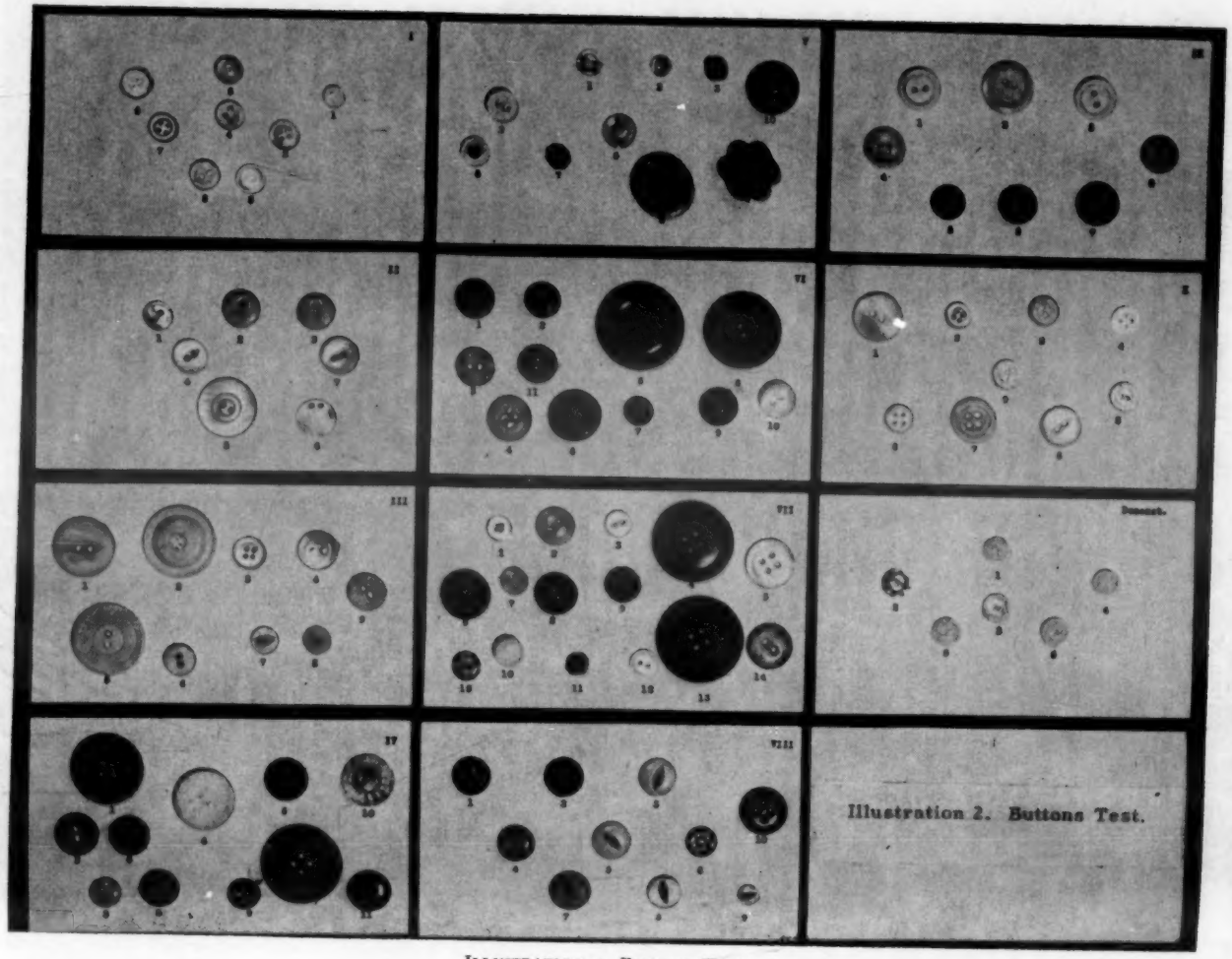

ILLUSTRATION 2. Buttons Test

which had the holes running perpendicularly to the face. On this card an alternate answer was also possible. On one of the buttons the material consisted of pearl and metal, making possible an alternate correct response of being made of two different kinds of material, whereas the others were all of one material. On card No. 6 all buttons were monochromatic except one which was two-colored. On card No. 7 all the buttons were round except one which was somewhat square shaped. On card No. 8 all buttons except one had grooves or indentations for the thread, probably to prevent the thread wearing out through friction. One button lacked this groove. On card No. 9 all buttons except one had some sort of rim extending around the outer periphery. On card No. 10 all buttons except one had a shiny appearance due to the fact that they are either simulated or real mother of pearl, whereas, the different button was somewhat dull, and made of a different material, perhaps palm nut.

The instructions for this test were as follows: "I have here some cards with various kinds of buttons on them. What I want you to do is to look at these buttons and tell me which button does not belong in the group. Look at this card (presenting demonstration card). All of these buttons except one are alike in a certain way." Point to the one which is different. After this was done say, "Yes, that is right, remember that all the buttons are alike except one, in a certain way. I want you to tell me which one is different." (Present first card.) "Which one is different here?" After the button was chosen by the subject, he was asked, "And how are all the other buttons alike?" The first time an incorrect response was made on a card, the error was shown to the subject by pointing out how the remainder of the buttons were not alike in respect to the way he had made his selection. He was then allowed to continue with the same card until he achieved the correct solution or was unable to see a solution, whichever the case might be. This procedure was followed for each additional card. However, only one such correction was made for each card.

One point was given for the correct selection of the "different" button and one credit was given for an adequate verbalization of the way in which all the other buttons were alike. In some cases where the subject was certain that he had selected the right button, but was unable to give an adequate explanation of the similarity among the others, he was given one credit. The total score was a summation of the scores of the individual cards.

This test was included in the battery to provide an analogous test to Tests No. 10 and 12 as mentioned above. In general, it was felt to be a test of concept ability using fairly common objects. On the basis of the preliminary testing, the test was found to be one which evoked considerable interest and left a feeling of general good will toward the test battery as a whole. It was a relatively easy test compared to some others in this battery and of genuine interest to practically all subjects.

ILLUSTRATION 2. Buttons Test

A. Statistical Treatment of the Results

An investigation centering around the area of reasoning was undertaken. A preliminary selection of tests purportedly measuring this area was made. These tests were modified if necessary, for feasibility of administration to a high-school population. On the basis of this investigation tests were modified for the study. One hundred subjects were selected in the junior class of a medium-sized high school following administration of the Primary Mental Abilities test (SRA) to the entire class. The remaining sixteen tests were administered to the one-hundred subjects selected as an experimental group. Testing covered a period of approximately two months.

Pearsonian coefficients of correlation were calculated (see Table 1). Factoring was done using the group centroid method of Thurstone (32, pp. 149 ff.). The method outlined on these pages, including checks, etc. was followed closely. After the seventh factor was extracted, the residuals appeared to be too low to merit another extraction. In order to provide rotational freedom and insure the extraction of all possible meaningful factors an additional factor was extracted. This factor could be discarded later if necessary without causing too much computational adjustment and was felt to allow convergence of solution somewhat earlier than otherwise would have been the case. In general, in the author's opinion, in an exploratory study as this the extraction of all pos-

TABLE 1

Intercorrelations of Tests

	M	σ	PMA V	PMA S	PMA R	PMA N	PMA W	Wech II Similarities	Wech II Comprehension	Binet Vocabulary	Judgment	Word Sorting	Concept Fluency	Opposites	Gottschaldt Figures	McGill Picture Anomalies	Picture Titles	Rorschach Z Beck	Arthur Stencil Design	Wech II Block Design	Holzopple C.F.	Hanfmann-Kasanin C.F.	Buttons
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	24.10	8.490																					
2	18.45	13.102	.281																				
3	14.18	6.160	.529	.279																			
4	20.11	8.212	.383	.136	.362																		
5	22.93	10.479	.467	.291	.427	.410														PE_r = 0–20 = .067			
6	11.07	2.850	.441	.301	.306	.240	.343														20 = .065		
7	10.43	1.940	.173	.138	.149	−.024	.119	.210													30 = .061		
8	5.54	2.733	.632	.459	.473	.157	.250	.477	.288												40 = .057		
9	30.72	6.888	.193	.295	.290	−.009	.067	.014	.124	.269											50 = .051		
10	10.61	3.501	.480	.305	.450	.136	.230	.354	.232	.581	.372										60 = .043		
11	18.78	11.055	.174	.227	.170	.028	.253	.189	.205	.346	.096	.137									70 = .039		
12	17.63	4.379	.303	.348	.472	.057	.148	.328	.252	.435	.259	.345	.240										
13	6.72	4.398	.259	.207	.186	.040	.193	.191	.334	.240	.142	.238	.050	.200									
14	15.88	1.986	.268	.164	.105	−.083	.001	.205	.216	.377	.167	.183	.220	.156	.199								
15	14.85	3.254	.179	.277	.188	.031	.027	.230	.336	.317	.226	.223	.284	.130	.153	.342							
16	16.61	9.625	.148	.033	.074	−.052	−.018	−.048	.182	.148	.090	.191	.234	.090	.247	−.013	.240						
17	10.26	5.240	.250	.485	.296	.112	.228	.301	.139	.471	.291	.306	.242	.358	.444	.300	.223	.050					
18	23.92	7.953	.279	.445	.463	.148	.263	.264	.245	.306	.242	.229	.098	.324	.447	.155	.210	.104	.577				
19	20.83	3.023	.109	.291	.209	.125	.081	.261	.270	.195	.322	.296	.080	.186	.151	.067	.170	−.028	.228	.273			
20	19.63	6.640	.196	.246	.316	.087	.242	.182	.176	.280	.292	.276	.190	.284	.226	.002	.103	.015	.336	.262	.331		
21	13.66	2.899	.270	.360	.202	.131	.129	.275	.154	.442	.317	.369	.116	.284	.156	.201	.247	.068	.317	.257	.183	.113	

TABLE 2

CENTROID FACTOR LOADINGS (TABLE A)

	I	II	III	IV	V	VI	VII	VIII	h_8^2
1	645*	−434	−124	153	027	138	179	−112	7075
2	594	055	127	−184	−261	−143	−032	099	5052
3	629	−307	245	148	146	−173	−028	−152	6470
4	274	−421	217	240	−144	077	−076	101	3996
5	448	−306	200	307	−164	128	−144	−111	5049
6	544	−262	−085	−087	−205	212	−194	170	5329
7	413	232	−198	156	165	130	−170	126	3768
8	755	−173	−306	−196	−030	026	089	−137	7603
9	430	180	099	−244	253	−186	062	−058	3925
10	633	−172	−117	−183	307	034	097	070	5872
11	381	103	−200	085	−142	−092	−215	−276	3540
12	551	−040	040	−132	133	−180	−093	−118	3969
13	461	262	061	186	095	269	279	108	4904
14	350	139	−328	−091	−167	076	138	−179	3424
15	435	285	−354	097	−120	−144	−126	122	4711
16	194	160	−227	326	220	−229	164	106	3600
17	634	245	233	−146	−270	085	251	−102	6911
18	599	203	352	131	−096	−046	160	110	5901
19	401	160	155	−175	139	127	−257	219	3905
20	436	091	287	−082	189	126	−202	−174	4101
21	489	−031	−084	−241	−091	−130	133	173	3780

* Decimal points are omitted.

sible factors is justifiable. The unrotated factor matrix is shown in Table 2.

The method of extended vectors of Thurstone (32, pp. 225 ff.) was used for rotation because this method offers a rapidly convergent solution. Thurstone maintains that this particular method is suitable for batteries which show a positive manifold or positive correlations. It is not suitable for bipolar factors. Either an oblique or orthogonal solution is possible. Simple structure expressed in terms of minimizing as many loadings of tests on factors as possible was one goal of rotation. In general, the most important criterion used in the preliminary rotations was psychological meaning. For this purpose, the tests in the battery which had been found to have high loadings in certain factors were not allowed to influence rotations unduly, although it was felt that certain of the tests used in the battery for purposes of identification would have high weights in this analysis. In general, the tests showed good structure on the plots from about the third rotation and the remainder of the rotations was made to reduce angles, and obtain simple structure. One of the rotated factors was not

TABLE 3

COSINES OF ANGLES BETWEEN PLANES

	A	S	J	Pr	C	Rp	B	Fx
A	1.000							
S	−.012	.999						
J	.058	.042	1.000					
Pr	.195	.115	−.029	1.000				
C	.000	.003	−.130	−.009	1.000			
Rp	.030	.014	−.119	−.014	−.026	1.000		
B	−.014	−.031	.025	−.029	.042	.063	1.000	
Fx	−.032	.036	.033	.022	.002	.006	.010	.999

TABLE 4
CUMULATIVE TRANSFORMATION MATRIX

	A	S	J	Pr	C	Rp	B	Fx
I	397*	240	459	417	345	336	240	365
II	−792	347	−151	063	012	427	147	107
III	130	709	−053	012	−022	−368	440	−208
IV	229	004	−464	843	−143	057	−129	−194
V	−024	−053	311	011	203	159	−024	−870
VI	174	465	−304	−147	340	205	−672	−006
VII	−321	123	377	295	180	−690	−402	128
VIII	−108	−290	−466	−047	819	−152	311	−002

* Decimal points are omitted.

satisfactorily meaningful in terms of the criteria heretofore mentioned and consequently on final rotations it was necessary more or less to ignore this plane in order to achieve satisfactory solutions with the other factors. In general, simple structure was not complete in the battery as can be seen from an examination of Table 5. However, with some of the factors it is reasonably achieved. Thurstone points out that in experimental factor studies complete simple structure is rarely attained (32, p. 256). In general, the solution of the problem is orthogonal as can be seen from the angles in Tables 3. The transformation matrix is shown in Table 4.

TABLE 5
INTERPRETATION MATRIX (Λ)

Test	A	S	J	Pr	C	Rp	B	Fx
1	597*	034	383	407	191	012	−183	184
2	148	166	247	102	203	102	376	456
3	550	171	412	397	028	029	271	−012
4	556	079	−077	253	116	−124	082	078
5	601	227	006	371	−004	080	052	153
6	478	015	054	037	334	216	059	357
7	050	034	−018	243	278	471	042	020
8	344	−032	560	164	200	241	−049	396
9	−067	159	464	037	135	139	281	−008
10	290	000	497	121	394	150	051	017
11	149	004	162	196	−200	399	073	272
12	233	078	434	122	060	212	279	088
13	−031	369	085	405	384	116	−110	099
14	−042	012	260	110	013	213	−186	391
15	−072	−159	032	254	151	423	166	333
16	−114	−186	076	443	110	094	044	−097
17	006	516	342	223	173	040	126	504
18	084	436	158	425	266	014	315	246
19	090	218	039	−072	366	328	293	010
20	221	414	235	054	059	297	193	−066
21	071	−043	326	047	308	009	197	336

* Decimal points are omitted.

Chapter III

SUMMARY OF THE FINDINGS

A. Factor A

It is felt that this factor can be best described as a learned activity factor that includes all those things that we ordinarily call *"academic"* in nature. Its primary expression in the test is through verbal material, but this fact is not surprising when we consider that most of our learning takes place through use of and by means of verbal processes. It involves the use of symbols and abstractions. In some respects it may represent the part of one's learning which is somewhat artificial or scholarly for use in school.

On factor A tests with substantial loadings are 1, 3, 4, 5, 6, 8, 10, 12, and 20 (see Table 5). This factor was most strongly identified in the Primary Mental Abilities portion of the battery. It was moderately strong in the adaptation of the Binet Vocabulary No. 8 and in the author's Word Sorting test No. 10. The highest weight is in the Word Fluency test of Thurstone and only slightly less strong in his Verbal test (Vocabulary). This at first suggests a *"Verbal"* factor until one attempts to explain the weight of the number and the reasoning test scores in Thurstone's tests. It would also seem to be a factor which is found almost exclusively in the type of activity required in a speeded paper and pencil test of the type used in this analysis, except for the fortunate circumstance that several of the individual tests have fairly substantial weights (No. 6, .478; No. 8, .344; No. 10, .290; No. 12, .333; and No. 20, .221).

This factor is quite similar to the "restrictive task" factor identified in Thurstone's analysis of perception (29) and is similar to the "central integrative" factor of Halstead (14) which he describes as the "well-organized habits of the individual." Halstead's belief that this factor has a high loading on so-called "psychometric intelligence" is thus justified if one confines oneself to paper and pencil tests and does not consider "psychometric" to include the individual tests in the repertoire of the clinician. Since Halstead derives his description from a wider range of material than was included in our test battery, he might have some justification for his somewhat broader factor name. In our battery the principle of parsimony and the lack of such range must limit our description to a so-called *"academic"* factor or learning factor. Certain of the tests which are quite similar in nature yield some support for the identification of this factor. For example, test No. 5, Word Fluency, in which the subject is required to name as many things as he can think of within a certain period of time that begin with the letter *s*, when compared with test No. 11, the author's naming test in which he is required to name red things, round things, etc. shows a rather substantial difference. In the case of the latter, the loading is only .149. According to some authorities both these tests should have loadings in and would primarily measure some sort of fluency factor. In terms of learning it is felt that initial letters are of considerably more fundamental nature than the enumeration of objects according to certain standards as required in test No. 11. The construction of dictionaries by alphabetical arrangement rather than by conceptual classifications is one illustration of this difference. Tests No. 19 and 20 show a somewhat less startling difference, but one which yields some slight corroborating evidence. The first test is a block-sorting test in which a number of variables are taken into consideration, while No 20 is a similar test with a single solution dependent upon the characteristics of height and width of the blocks. The way in which this test difference adds to the evidence is that one of the most usual methods by which objects are compared is by a criterion of size. The word "big" or the word "little" is felt to become a part of the young child's learned-response repertoire before such things as the other variables measured in test No. 19. In general, the tests which employ that which is new or unfamiliar or unusual have zero loadings on factor A. This is believed to be the reason why test No. 2 (the Space test) in the PMA series has relatively low loading, particularly when compared with the other subtests of this group test. Tests No. 7 and No. 9, both of which employ verbal material, are more difficult to explain, but it can be postulated that the requirements of these two tests are somewhat more discriminatory than is ordinarily needed in the course of one's educational history. Tests 13, 14, 15, and 16 all have slight negative weightings in this factor. In each case it can be argued that the material has little representation in terms of the psychological functions involved in learned activity. (It is of some interest to note that the *Z* score on the Rorschach has a weighting of −.114. The remainder of the tests, except for No. 20, were all relatively new tasks for the majority of subjects.

While this sort of a factor called *"academic"* in our analysis has been identified in other works as noted above, it is felt that the description herein given is well defined and identified by the tests in the battery. Subjecting this factor to our criteria for adequate factors suggests that this factor meets the standards satisfactorily. It should be of some interest to both clinicians and general psychologists as well as others interested in psychological functioning of human beings to be able to attain measures of the effects of one's learning experience. It might, if validated by further studies, be conceivable that if such a factor can be accurately identified, we could use it as one approach to study some of the shortcomings of our educational theory, and more particularly, our educational practice.

It is the opinion of the author that psychometrics must be broadened to measure other than the characteristics measured in the average group intelligence test. This may be difficult, but certainly the whole area of limitations of tests of particular types needs more airing. One has only to see people classified in terms of traditional tests as "borderline-feebleminded" functioning adequately in everyday social contacts to realize that tests sometimes do not tell the whole story. In our high-school population one of the surprises of the entire battery was to find that five persons had "quotients" of less than 70, the usual dividing line between feeblemindedness and normality. Many teachers would undoubtedly agree that they had "morons" in their classes, but they were probably employing the term somewhat less carefully than a psychologist should. One of the striking characteristics of some of these "borderline-feebleminded" subjects was that on some of the tests in the individual battery they achieved scores which were considerably closer to the mean than one would expect and in some cases were considerably above the average of the rest of their classmates. It is not felt that the author was fortunate to encounter a large group of idiot-savants among the hundred subjects used. It was felt that, instead, he was dealing primarily with academically handicapped people—in most cases probably poor readers. On the basis of social intelligence and other nonacademic activity, the "moron" group appeared to be relatively normal.

B. Factor S

This factor is likely most closely identical to some *"spatial"* factor in which the size, shape, or appearance of objects plays a predominant role. It is felt that this factor must be regarded as lacking entirely adequate psychological interpretation in terms of the particular tests employed.

Tests No. 18, 17, and 20 have rather substantial weights in this factor. Test No. 13 has a moderate weight and Tests 5 and 19 have slight weights. All these tests except No. 5 are tests in which concrete objects need to be manipulated. Only one test, test No. 5, is sufficiently different from the others to merit consideration, but this test has such a weak saturation (.227) that the rules of evidence rule out its having much weight. One rather remarkable weak test in this factor is the test employed in the PMA series to measure space. This suggests several alternatives, either the test does not measure space, or this factor is not a space factor, or it is not the kind of space factor that is measured by the other tests in this battery. The last alternative is believed to be the correct one. As mentioned above, the visual elements necessary to satisfactory performance on tests having substantial weights are primarily concerned with shape or size. In general, the very slight weight in No. 5 suggests that some people perform in this test by means of visual aids. This has some clinical validation in the behavior of subjects taking this particular test. One need not be a very keen observer to notice certain individuals performing on this test by use of sounds and one can actually see many students moving their lips and mouths to suggest that they are combining *s* words with various vowels to suggest to themselves words to be written. Similarly, one observes others looking around to get ideas from their immediate environment to aid them in the test. Similarly, one can easily observe a number of people in most groups taking this test who are "looking" but apparently not at anything in particular. Some even close their eyes, but by their expression, manner, and behavior suggest a visual activity. It is the last group which the author feels are achieving results in the test by the visual aid of seeing words beginning with *s* in their imagery. If the clinical analysis of this test performance can be validated, experimentally or otherwise, it would suggest a somewhat more psychologically oriented meaning for this factor, namely a *"visual imagery"* factor. As limned in this battery, however, we must on the evidence confine our interpretation to a spatial factor in which shape and size are of primary importance. Perhaps further exploration of this factor may divide people into some sort of dichotomy as that which the layman calls "concrete" thinkers and "abstract" thinkers, or it may lead to a division into groups such as auditory types versus visual types. Fortunately, the techniques of factor analysis do not make these groups antithetical but merely presuppose that they are uncorrelated, at least when

found in an orthogonal solution. Perhaps of some importance in identifying this particular factor are slight negative weights in Test 15 (—.159) and in Test 16 (—.186). In the latter tests, visual imagery should not be a handicap unless it is a precise sort of imagery. It does lend some support, however, to the size and shape aspect of the interpretation of this factor because one has to allow himself liberty in distortion of both these characteristics to achieve adequate concepts for better scores.

C. Factor J

According to the author's analysis it would seem most likely that this factor can be most adequately described as *"judgment."* This judgment finds its expression in the weighing of hypotheses, evaluation of thought materials, criticism, and discrimination. It finds its expression in a wide variety of tasks in which *control* of thinking is of primary importance. The latter suggests a possible relationship between this factor and the concept of maturity or wisdom or the things one gains as he gets older and wiser, perhaps what we call experience.

This factor has high weights in Tests 8, 3, 9, 10, and 12, with moderate weights in 1, 17, and 21. This is obviously some sort of factor revolving around the sphere of reasoning. An examination of these tests is at first suggestive of some sort of inductive factor since test No. 3 is theoretically an induction test as are tests No. 10, 12, and 21. Similarly, one could present a fairly strong case for deduction in terms of its expression in Test 8 (the Binet Vocabulary test), No. 9, and No. 1, because some analysts have felt that in these tests persons bring a wealth of material to bear upon solutions by a deductive principle (see French [10] for example). The hypothesis of induction is not sustained by the weights that tests have that are readily interpreted as deductive in nature and vice versa. From the standpoint of an analysis using this sort of technique neither induction nor deduction would appear to be very clearly established.

It is felt that this name *"judgment"* serves to explain better the differences between tests which are somewhat similar in character, since Test 8, the Binet Vocabulary test, has a higher weight than Test 1, the PMA vocabulary test. One could argue that in the somewhat freer situation of the former that the subject needs to exercise considerably more discrimination, control, etc. than he needs to in Test 1, in which he merely has to choose among several alternatives instead of the mass of material coming to consciousness in the former case. Of the three concept-formation tests devised by the author, Buttons has the lowest weight, which suggests that the fewer limitations imposed upon the reasoning process in this test caused it to be somewhat weaker than Tests 10 and 12. The greater latitude in administration might also have played a role. Tests 17 and 18 might also be cited as additional evidence. In the first one a trial and error performance generally has much less effect than a similar performance on Test 18, where such a procedure is frequently quite valuable in achieving a correct solution—at least for some subjects.

D. Factor Pr

It is apparent that these tests are ones which center around two foci, namely speed of operation and fluency of ideation. Considering the two together, one can describe this factor as mental *"productiveness."*

On this factor, tests No. 1, 3, 13, 16, and 18 had high loadings, while tests No. 4, 5, 7, 15, and 17 had moderate loadings. All of the Thurstone tests are represented here except No. 2. Just as the amount of flow through a pipe is dependent upon the size of the pipe and the pressure, similarly mental productiveness is dependent upon speed and upon a number of ideas which come to one's mind that pertain to the problem. In tests in which the time is unlimited, the weight on this factor would be primarily dependent upon one's wealth of ideas. In a limited time, however, this factor would depend largely upon the degree of speed with which ideas reaching a conscious threshold are produced. Tests No. 1, 3, 18, 4, 5, and 17 are examples of tests in which the speed element is of paramount importance. Tests 16, 7, and 15 are examples of tests in which the richness of ideas causes heavy weights. Test No. 13, it will be recalled, was not a speed test and the subject was given unlimited time in which to perform. However, the instructions for this test were such that in essence, this test became a speed test. It is of interest that, allowed one's own pace, this factor still emerges. It is rather remarkable that some of the tests which are definitely speed tests do not emerge in this factor. One wonders, for example, why test No. 2, the PMA Space test, does not have a high weighting, because the score in this test is dependent upon the number of figures correctly identified as unreversed in the period of time allowed. It is of some interest to note that in this test a total of 19 subjects received zero scores. This includes, of course, subjects

whose scores were negative in addition to those whose scores were actually zero. Since PMA standards were used in the scoring of this test, no attempt was made to evaluate scores below zero. In addition to these 19 subjects having zero scores, an additional 14 subjects had scores which were so low they might well be attributed to chance. This grouping on the test into two divisions, one of which scored significantly and a group which scored at a chance level, suggests a test which is perhaps too difficult or too crude in discriminatory power to be very useful. Part of the failure of this test to emerge with significant weight on this factor is perhaps due to some of these inherent weaknesses. Another test, in which speed was an element in scoring, that did not have weight on this factor was test No. 20. It was a rather common observation that among persons who worked rapidly on this test, the number of errors made appeared to be uncorrelated with the time, or perhaps was slightly negatively correlated. This test, however, does have more weight than its nearest likeness, No. 19, on this factor. Another test in which a time limit was used, not expressed in this factor is test No. 11, in which subjects were asked to name as many red things, etc. as they could think of in one minute. It was generally true that most subjects would exhaust their thoughts after thirty seconds on most of the items of this test. Yet, they were not given an unlimited time in which to perform as in tests No. 7, 15, and 16.[3] On the latter tests, more than one response is required except for some of the items in test No. 7. It was necessary for the subject in most cases to give these responses within a reasonable length of time. It was felt that in some cases subjects might conceivably have produced answers if they could have ruminated over the items, but the general pace of the tests suggested to them that they should not spend too long on individual items. This factor finds expression in a great variety of tasks, yet is uncorrelated with other factors. One might expect a factor in which speed is the most important element to be correlated positively with all other factors and lead to an oblique solution for the factor. A study of the angles (see Table 3) shows that while it has a slight positive correlation with our first two factors (*"academic"* and *"spatial"*), with some of the others it shows slight negative correlations. None of these correlations, however, is sufficiently high to suggest an oblique character for this factor.

The test battery in general was designed to permit emergence of possible speed factors. However, this was not the primary problem under consideration in the study, and, consequently, it did not emerge in our battery with as much clarity as in a study of this specific area by Davidson and Carroll (9). Their technique of using both speed and power scores for the same tests reveals this factor with more certainty. While our data do not approach this, it is felt that in general this factor is fairly well defined in the battery. As they point out, the implications of such a factor are many. Certainly we can no longer afford to make all tests speed tests or time-limit tests or power tests. In an adequate test battery, the common factor variance must not be seriously increased by such uniformity of testing procedure. In general, this factor seems to be relatively unchanged in shifting from one type of material to another. It is suggested that failure to take this factor into account has resulted in correlations among tests that are difficult to explain because they possess a spurious element, especially when speed is not a defined factor.

[3] On a majority of the items in No. 7, however, two responses are necessary for maximum credit. On No. 15 and 16 three responses are required.

E. Factor C

With the exception of Test 13, these tests are ones which were supposedly measures of concept formation—the ability to grasp the essential similarity between discrete variables, to abstract them, and to generalize them. The idea of finding the rule or of reasoning from the specific to the general as induction, is apparently a similar way of saying the same thing. We have named this parameter the *"concept"* factor.

This factor was rather weakly identified in the test battery. Tests 6, 10, 13, 19, and 21 have moderate weights. Test No. 3 and test No. 12 are missing members of the group that was hypothesized as measuring this area. This ability seems to be, in our analysis, limited to types of tests in which choice is unlimited. This might serve to explain the lack of weight of test No. 3 and the failure of test No. 20 to match test No. 19 in loading. Test No. 13 is, in a crude sort of way, a test of finding similarities and might be on a superficial basis called a test of abstraction of a common element in spite of a mass of superfluous perceptual material. Test No. 12, which was hypothesized as a concept-formation test, has the characteristic of giving answers in terms of differences and this may be the reason why this test does not have much weight in this parameter. It is perhaps of some significance that test No. 11, a supposedly deductive test, had a negative weight of —.200.

F. Factor Rp

This is operationally described in terms of willingness to respond where it was largely up to the individual to make his own rules of suitability in a deductive fashion. It might imply some suggestibility or overanxiousness to please the examiner, although this was not felt to be too important. It is also notable for a lack of a critical element in thinking. We call it *"responsiveness."*

This factor was neither clearly nor strongly identified in the battery, but the weights of the tests are rather strongly indicative of good psychological meaning for this factor. Tests No. 7, 11, and 15 are weighted heavily in this factor, while tests No. 6, 8, 12, 14, and 19 are significant. A review of these tests is suggestive of a factor which is perhaps only partly cognitive and perhaps partly temperamental in character. Recall of the directions and characteristics of Tests 19 and 20 will serve as some corroborating evidence for description of this factor as *"responsiveness"* since Test 19 allows considerably more freedom of choice and initiative on the part of S.

G. Factor B

This factor is too weakly identified in the battery to merit interpretation. It appears to be rather complex and is very slightly suggestive of some sort of analysis-synthesis dichotomy since its structure is somewhat bipolar. During the rotational process this factor caused a great deal of difficulty by its vagaries and it was not until it was ignored that rotation toward simple structure and orthogonality was easily achieved.

H. Factor Fx

This would appear to be a sort of ability to change one's mental set, or plasticity, or mental flexibility, or ability to try a new tack. It is felt that mental *"flexibility"* best subsumes this factor.

This factor is expressed in tests No. 2, 6, 8, 14, 15, 17, and 21. It was no surprise that test No. 17 emerged as having highest weight in this factor. After watching a great number of subjects attempt to reproduce the designs presented to them in this test in a reversed manner in which white served as a ground function rather than the figure, it was anticipated as some sort of inability to shift perceptually in terms of figure and ground relationships. Test No. 2, Space, as noted before, showed a rather remarkable dichotomy of distribution in terms of scores. From the author's own introspective analysis a great deal of shift is necessary to perform adequately on this test. The shift here, however, is not in terms of figure-ground relationships but is expressed in terms of rotational space. Somehow or other, the subject who performs well in this test transposes the figures among the answers to see whether or not they are congruent with the given figures. The great distress that a number of subjects taking this test always show is indicative of their inability to make this transfer. With this factor expressed in terms of these two tests alone one might assume some sort of a perceptual breakdown or analysis involved. Fortunately, however, it finds some expression in a variety of other tests, such as test No. 6, Similarities, and test No. 8, Binet Vocabulary, tests in which a considerable verbal element is involved. It is felt that the weight of test No. 6 is primarily due to the ability to eliminate superficial elements of similarity and to choose more fundamental (and better scoring analyses) on the early items of this test. On the later items, the inability to think in terms of similarities rather than differences is one important cause of failure. It is very frequently encountered in using a test of this particular type that dull subjects will persist in giving answers such as "they are not alike" and regard such an answer as entirely adequate. Such a performance is felt to be somewhat characteristic of people of average ability, too. Although they realize that some similarity might exist, they are so overwhelmed by the number of differences that come to mind that they are incapable of correct performance. The weight of test No. 8 is somewhat more difficult to explain. It is felt to have weight because of a difficulty in bringing to consciousness the essence of a definition with sufficient clarity to achieve a satisfactory response. It is generally recognized that multiple-choice vocabulary tests, as test No. 1, covering a similar range of difficulty, would likely result in higher scores than a test in which definition is required. The reader with experience in any kinds of tests will realize on the basis of his own experience that completion items are usually more difficult than choice items. It is felt that the reason for this is that possible answers are considerably more numerous in the free situation that exists in completion and the difficulty of settling on any one answer is consequently greater. Test No. 14 is a recognition of absurdity in pictures. The weighting of this test is suggestive of a similarity between this fac-

tor and that quality of individuals called a sense of humor. One of the characteristics of jokes is perhaps the sudden shift of set usually involved. Incongruity may be the essence of humor and perhaps the person who possesses a good sense of humor is one who is capable of recognizing absurdity. If this is so, it is notable that one dimension of intelligence or reasoning is necessary in order to get the point of a joke. Test No. 15, Picture Titles, has weight in this factor because of the necessity of the subject to think of some other way of describing the picture other than in the terms that he has previously employed. In general, one response was easy for all subjects; some people experienced difficulty in giving two responses; a considerable portion of the group was unable to give three responses, particularly for some pictures. The remaining test that had some weight was the Buttons test. In this test a different response is required for each set of cards and a new and different method of grouping must be employed. The persistence of memory of previous solutions was generally felt to be a handicap for a later performance. Those who gave satisfactory responses on some of the easier items on this test and were unable to recognize that it would be unlikely that the same principle would be used later gave less satisfactory responses. On the other hand, those who looked upon each item as a new challenge, with their minds relatively free of prejudices, apparently performed most adequately. It is of some interest to note that test No. 21 has greater weight than tests No. 10 and 12. The reason for this is that No. 10 and 12 place somewhat less emphasis and are somewhat less broad in the sense of the number of things to be taken into consideration than test No. 21. Test No. 16 has no weight in this factor because of the fact that shifts in set resulted, as often as not, in lowering the *Z* score. Most frequently the shift took place in terms of location of response rather than in terms of organizational ability involved in giving a new response having a high *Z* loading.

Chapter IV

A DISCUSSION OF TEST COMPOSITION

One aspect of the problem in this study was to discover the factor composition of various mental tests, particularly those which had not yet been subjected to factor analysis.

In general, the five Thurstone tests in the battery do not in our analysis emerge as "pure" measures. This is because, it is felt, the battery used did not have enough tests of a similar nature which could combine to allow for emergence of the Thurstone factors. However, they did show significant weights on some of the factors found in this analysis. It would thus appear that Thurstone's belief that tests rotated to simple structure in one test battery would retain their factor composition in other test batteries has not been upheld in this analysis. In the author's opinion, part of the failure of these tests was produced by limitations of the method of test administration, and the type and characteristics of the test itself.

Test No. 1 has significant weights on the *"academic"* factor, the *"judgment"* factor and the *"productiveness"* factor. In terms of the parameters used in this test it can be described as a measure of learned activity that requires discrimination and evaluation in its solution. Test No. 2, the Space test, in this battery was quite complex factorially and was expressed in the battery with moderate weight in the *"judgment"* area and with strong loading in mental *"flexibility."* It also had considerable strength in the unknown and unnamed Factor B and consequently might be assumed to have certain unknown loading. The Thurstone Reasoning test, No. 3, has high loadings in the *"academic"* factor, in the *"judgment"* factor, in the *"productiveness"* factor, and moderate weight in the unnamed factor. In this respect it would appear to be quite similar to test No. 1 in terms of what it measures. Test No. 4, Number, has a high loading on only one

factor, namely the *"academic,"* and a moderate weight on the *"productiveness"* factor. Test No. 5, Word Fluency, measured *"academic"* ability, *"productiveness,"* and had slight weighting in the *"spatial"* factor of this study.

The Wechsler Similarities test was weighted with our *"academic"* factor, with the mental *"flexibility"* factor, and the *"concept"* factor. Test No. 7, Comprehension, was most heavily weighted in our *"responsiveness"* factor. It had moderate loading also in the *"concept"* factor. Test No. 8, the adaptation of the Binet Vocabulary, had expression in the *"academic"* factor, the *"judgment"* factor, and the *"flexibility"* factor. Test No. 9, Judgment, only found expression in the *"judgment"* area, although some of its variability was unaccounted for by its inclusion in the unnamed factor. Test No. 10, Word Sorting, was a fair measure of *"judgment,"* was among the strongest tests in the *"concept"* factor, and had modest loading in the *"academic"* factor. Test No. 11, Naming test, had its chief expression in the *"responsiveness"* factor. (It had a weighting of −.200 on the *"concept"* factor.) Test No. 12, Opposites test, is apparently a moderately good measure of *"judgment,"* with slight weighting in the *"academic," "responsiveness,"* and the unknown parameter. Test No. 13, Gottschaldt Figures, was a good measure of *"productiveness,"* and of *"concept"* ability. Test No. 14, Picture Anomalies, measured mental *"flexibility"* and to a lesser extent *"judgment."* Test No. 15, Picture Titles, measures *"productiveness"* to a slight extent and *"responsiveness"* and *"flexibility"* to a somewhat greater extent. Test No. 16, Beck's Rorschach *Z* score, appeared to be a relatively pure measure of *"productiveness."* Test No. 17, the Stencil Design test, turned out to be complex factorially, having important weights on the *"spatial"* factor, on the *"judgment,"* and on the *"flexibility"* factor, as well as a lesser weight on the *"productiveness"* factor. Test No. 18, Block Designs, measures *"spatial," "productiveness,"* and to a lesser extent *"concept"* ability and *"flexibility."* Some of its variance was unaccounted for by a moderate expression in Factor B. The Holsopple Concept Formation Test measures *"concept"* ability and *"responsiveness."* The Hanfmann-Kasanin test, No. 20, measures our *"spatial"* factor and to a lesser extent, *"responsiveness"* and *"judgment."* The Buttons test appears to be moderately good measure of *"judgment," "concept"* ability, and *"flexibility."*

In general, the predictions made concerning the factor composition of new tests were sustained. Few tests in the battery emerge as pure measures of any factor, except the Rorschach *Z* score, Picture Titles, and Judgment. This fact suggests two things: first, pure measures are difficult to construct, and second, our factors enter into a broad range of human behavior as abstracted in our tests. In general, they lose some value as differential diagnostic instruments because of this fact. The author makes no apology, however, for his inability to predict or limit his tests to a particular factor since the published and widely used tests of Thurstone fail to meet a similar criterion in this experiment.

If, as is maintained in this study, factors are primarily parsimonious descriptive phrases of "vectors of the mind," it is questionable whether the hope or attempt of some factor analysts to achieve a set of tests which are pure measures of particular factors will likely be achieved. Perhaps our approach to this dilemma needs modification toward

the somewhat old-fashioned statistical technique of multiple regression.

A second alternative is to attempt to discover factors which have direct application for particular jobs that the psychologist must perform. This technique has been applied in a number of attempts to make up a suitable measuring and prediction device for vocational choice. In the author's opinion attempts to apply the findings of one area to other areas of interest to psychologists will prove either futile or inefficient. The failure of factorized tests to explain the neurotic, the psychotic, and organic behavior disturbances is perhaps due to the fact that they are not specifically designed for this purpose. It is not maintained that other tests of intelligence do a better job in this respect. It is sometimes distressing to the psychologist (but undoubtedly cheering to the patient!) to find that very gross destruction of cerebral tissues fails to change scores on tests. It is especially disconcerting when one observes the behavioral changes that take place in such cases and yet is unable adequately to define them in test terms. The failure to place one's finger upon the exact nature of changes and/or abnormalities of behavior reflects somewhat on our perspicacity as students of behavior.

Another solution to the general problem of prediction might conceivably lie in a direction of test design, particularly by paying somewhat closer attention to the administrative characteristics of tests themselves. For example, we might need to study the directions of tests in an experimental manner just as we study the test content. They may play a considerably more important role than we have realized. The findings of this experiment suggest that they aid in determining factor composition. Instructions to work rapidly have striking effects, for example. It would seem feasible to combine scores. For example, one test might give a part score toward a *"fluency"* factor, etc. and another test would yield part scores in these and other factors. By a method of statistical combinations the various weights could then be combined to give a more reliable measure of the parameters which have been found to have some applicability to a particular problem. It is conceivable that different combinations of the same scores might yield results useful in other areas. Again it is felt that the prescription will have to be specific. The hope of finding tests which measure only one factor and no other is believed to be somewhat more difficult than is commonly now realized, especially when one attempts to deal in the realm of the psychology of the individual. The great variety of approaches different individuals use in the same tests suggests an important difficulty. One only needs to discuss with ten different individuals how they go about deriving solutions to a letter-series test to discover that for some the solution is largely intuitive. For others it is a primary inductive task involving distance, while others see it has certain rhythmic characteristics and solve it that way. For perhaps one member of the group it might be a test of number series. Unless evidence is gleaned from a wide variety of sources, it is possible that pure measures of a certain factor might have vastly different implications from one individual to another. A variety of sources of weights toward a final factor score might minimize this difficulty.

It is difficult to see how some of the factors identified in this analysis can be purified. For example, it is difficult to imagine how the aspects of reasoning

herein described can be abstracted from the contents of the tests themselves without disturbing the meaning of the factor. Reasoning in general must always be *about* certain things and is carried on *by means of* certain symbolic processes. It is difficult to see how one can construct a reasoning test that does not use verbal material, perceptual material, or concrete objects. Depending upon the type of material used, the test might then contain factors in addition to those involved in reasoning.

CHAPTER V

CLINICAL VALIDATION OF FACTORS

DURING testing of some individuals rather extensive notes were made, especially when the subject exhibited unusual test behavior as in cases of extreme difficulty or when he gave evidence of particular ability. This was done with two purposes in mind. First, it offered a means whereby some direction could be given in the initial rotations to eliminate unnecessary work. A forecast of eventual factor composition was made on the basis of the clinical evaluation. The number of rotations necessary (twelve) is indicative that this technique was not without some merit. The second reason why notes were kept was an attempt to give some means of validating the factors found in terms of behavior of individuals since it was felt that this might serve as corroborating data for the statistical procedures. If this sort of empirical validation could be made, it was felt that it would insure that the factors found would be psychologically meaningful, and perhaps more important, that they would be useful in the sense that they could account for deviations in performance. It is generally believed that scatter in test performance is primarily due to the differential abilities in factors which various individuals possess. It would seem that the converse of this—that individual performance could be accounted for by variance in factor ability—should also be true.

From the completed tabulation, the scores of six individuals were selected without knowledge of meaning of the scores by a psychologically unsophisticated person. These six subjects then served as a test group to see if there was merit in this belief.

The scores of these individuals on all 21 tests were transformed to z scores in order to make them comparable. These scores are listed in Table 6.

The first subject selected had three scores which differed by 1.5 standard deviations from his average z score of +.31. These are respectively Test 3 (—2 standard deviations), Test 7 (+3 standard deviations), and Test 13 (+1.5 standard deviation). Test 3 was a test which measured the *"academic"* factor, the *"productiveness"* factor, and the *"judgment"* factor. In this particular case, other scores measuring the *"academic"* factor are somewhat depressed as are the tests in the *"productiveness"* factor, except for those with unlimited time. Scores on the *"judgment"* factor are inconclusive here, but there is sufficient evidence on the basis of the other two factors to account for the majority of the variability. Test 7, according to our analysis, measures primarily *"responsiveness"* with a moderate weight in *"productiveness."* Since scores on the *"productiveness"* factor are somewhat depressed, the variability here cannot be accounted for by this factor. On the *"responsiveness"* factor, Tests 11 and 15 are good measures. The average of these two tests is somewhat above the average standard score for all tests and probably plays an important role in its high weight. However, it does not sufficiently account for the extreme variation here. Test 13 was a measure of our *"spatial"* factor, our *"productiveness"* factor, and of *"concept"* formation. Its elevation can be as-

TABLE

RAW SCORES AND z SCORES

Tests

Subject	Type of Score	1	2	3	4	5	6	7	8	9	10
I	Raw Score	21	17	3	28	17	10	17	4	35	10
	z Score	− .37	−.11	−1.82	+ .96	− .57	− .38	+3.39	− .56	+ .62	− .17
II	Raw Score	16	24	9	13	0	8	9	2	40	8
	z Score	− .95	+.42	− .84	− .87	−2.19	−1.08	− .74	−1.30	+1.35	− .75
III	Raw Score	20	21	3	17	8	11	7	5	37	6
	z Score	− .48	+.19	−1.82	− .38	−1.43	− .03	−1.77	− .20	+ .91	−1.32
IV	Raw Score	13	23	9	8	10	11	10	3	35	4
	z Score	−1.31	+.35	− .84	−1.48	− .38	− .03	− .22	− .93	+ .62	−1.89
V	Raw Score	9	6	11	21	5	4	7	3	30	8
	z Score	−1.78	−.94	− .52	+ .11	−1.71	−2.48	−1.77	− .93	− .11	− .75
VI	Raw Score	33	9	4	24	13	8	9	3	24	8
	z Score	+1.05	−.72	−1.65	+ .47	− .95	−1.08	− .74	− .93	− .98	− .75

sumed to be due to the influence of the space element since other space tests are higher than average. However, concept formation scores are inconclusive.

In Subject 2, two scores of unusual character are present. Test 5 differs from the average z score by more than −1.5 standard deviation, while Test 9 is nearly +2 standard deviations above the average. Since in our analysis, Test 9 is a relatively pure measure of our *"judgment"* factor, its variability must be assumed to be due to presence to a high degree of such an ability. However, practically all of the other tests with high loadings on judgment are depressed. Therefore, we must regard this assumption as being disproven. Test 5, on the other hand, measures *"productiveness"* and the *"academic"* factor. In general, other tests measuring these factors are similarly depressed and it would seem to justify attributing the variability to these parameters.

Subject No. 3 has a score on Test 13 which is +2 standard deviations above the average. Since Test 13 measures *"concept"* formation, *"spatial"* and mental *"flexibility,"* its elevation would have to be attributed to these factors. Examining them in turn, it would appear that the spatial factor could account for a considerable portion of the variance, with some support from the mental *"flexibility"* factor. Tests of *"concept"* ability do not apparently contribute to the explanation for this anomalous score since they are ambiguous.

Subject No. 4 has one score on Test 15 at a point +2 standard deviations higher than the average of his other scores. Test 15, it will be recalled, measures *"responsiveness"* and *"flexibility."* Since other *"responsiveness"* tests are slightly depressed, the variability here must be attributed to the presence of ability to shift or the mental *"flexibility"* factor. The standard scores on other tests measuring this factor are considerably above average, so this would seem justifiable.

Subject No. 5 shows one score almost −2 standard deviations below his average. This is his Test 6 score. Test 6, it will be recalled, measures our *"academic"* factor, *"concept"* ability, and had a moderate weight in an unidentified factor. Other scores having weights in our *"academic"* factor are on the average considerably depressed, while tests having weight on the *"concept"* factor are slightly depressed. The combination of the two might well account for the variance.

Subject No. 6. The score on Test 1 is more than +1.5 standard deviations above average, while the score on Test No. 12 is more than −1.5 standard deviations below average. In general, these two tests measure different abilities, according to our analysis. Test 1 measures the *"academic"* factor, the *"productiveness"* factor, and to a lesser extent, the *"judgment"* factor. Support for the high score in Test 1 is not only lacking but in general is negative. The scores on other tests of a similar factor composition are in general depressed. Test 12, which measures judgment, is depressed as are practically all other tests measuring this variable.

In general, it would appear that on the basis of these particular cases, the evidence for clinical usefulness of our factors is slightly positive. In several cases, these factors fail to explain extreme deviations in scores. Out of the ten scores selected among the six individuals, six were readily explained, two were ambiguous, and two were contrary to what one would expect in view of other tests measuring similar factors. It is felt that

6
OF SELECTED SUBJECTS

11	12	13	14	15	16	17	18	19	20	21	M
19 + .02	18 + .08	15 +1.88	15 − .44	19 +1.28	34 +1.81	8 − .43	26 + .26	24 +1.05	19 − .10	14 + .12	+.31
9 − .84	12 −1.29	3 − .85	16 + .06	14 − .26	24 + .77	8 − .43	20 − .49	17 −1.27	19 − .10	9 −1.61	−.63
4 −1.27	18 + .08	15 +1.88	17 + .56	9 +1.80	13 − .38	13 + .52	30 + .76	23 + .72	12 −1.15	16 + .81	−.26
11 − .67	16 − .37	9 + .52	15 − .44	21 +1.89	32 +1.60	7 − .62	23 − .12	16 −1.60	8 −1.75	16 + .81	−.33
17 − .15	16 − .37	7 + .06	12 −1.95	13 − .57	27 +1.08	6 − .81	16 −1.00	23 + .72	27 +1.11	12 − .57	−.64
7 −1.01	7 −2.43	12 +1.20	16 + .06	17 + .66	17 + .04	3 −1.39	20 − .49	15 −1.93	10 −1.45	12 − .57	−.65

the ambiguity of this procedure merits further investigation, but it has not been proven that our factors are particularly useful in this respect. However, no means of comparison of the results with findings by other investigators are at present available.

CHAPTER VI

SUMMARY AND CONCLUSIONS

AN EXPLORATORY factor study in the general area of reasoning ability was undertaken to answer a number of questions, namely:

1. Can hypothesized reasoning factors be assigned to categories similar to (*a*) steps in a reasoning process, or (*b*) logical categories as induction, deduction, or (*c*) such descriptive terms as concept-formation ability, problem-solving ability, and analyzing and synthesizing ability?

2. What is the factor composition of certain tests which purport to measure reasoning or certain aspects of it?

Another aim of the study in general was to attempt to find factors which would have more psychological meaning in the sense that they would serve to explain a greater variety of human intellectual behavior, and would be more readily translatable into the descriptive concepts frequently used by clinicians. The position was taken that factors described in terms of test materials involved were inadequate and not sufficiently abstracted to be practical or useful except in limited circumstances. To this purpose, the statistical treatment of the data was always made subservient to meaning or usefulness. Two works by Meili and Halstead were especially recognized as excellent initial efforts in this direction.

A battery of tests was devised. They included the five subtests of the Science Research Associates version of the Primary Mental Abilities Test of Thurstone, which were administered in group form, and 16 individually administered tests. The five factors measured by the subtests of the former are: 1. Verbal, 2. Space, 3. Reasoning, 4. Number, 5. Word Fluency.

The remaining tests were:

6. Wechsler Form II, Similarities
7. Wechsler Form II, Comprehension
8. An abridged Stanford-Binet Vocabulary test
9. An abridged Test of Practical Judgment

adapted by the author from Cardall's test of the same name

10. A Word Sorting test devised by the author
11. A Naming test devised by the author
12. An Opposites test devised by the author
13. A Gottschaldt Figures test adapted for individual administration from the Thurstone Research Edition
14. The McGill Picture Anomalies Test, Form N, short test
15. A Picture Titles test devised by the author
16. Beck's *Z* score based on three Rorschach cards
17. An abridged version of the Arthur Stencil Design Test
18. The Wechsler Form II Block Design test
19. A modification of the Holsopple Concept Formation Test
20. A modification of the Hanfmann-Kasanin Concept Formation Test
21. A Buttons test devised by the author

The Primary Mental Abilities test, a group test, was administered to the members of the junior class in a medium-sized high school in central Pennsylvania. From this group, a random selection was made of 100 subjects who were given the remaining tests in the battery. Testing was conducted in the school in a suitable location during a period of approximately two months immediately succeeding the group tests.

From a correlation matrix, eight factors were extracted employing the group-centroid method as outlined by Thurstone. Rotation of these factors was made using the method of extended vectors as outlined by Thurstone.

The rotations, primarily determined by psychological meaning and usefulness, were made toward simple structure and orthogonal axes.

Seven factors were identified as a result of these rotations. A remaining factor was not capable of being interpreted. The first factor, A, was called an *"academic"* or *learned activity* factor. It found its expression in the test battery in vocabulary tests, number-computation tests, a word-fluency test, a reasoning test, and others. This factor was believed to be similar to a "restrictive task" factor of Thurstone and a "central integrative field factor" found in Halstead's work.

The second factor, S, was named a *"spatial"* factor having its chief expression in the ability to recognize size, shape, or congruence. There is suggestion but lack of sufficient evidence to identify this factor positively as some sort of visual-imagery factor. It was found in the battery in tests of Block Designs, Stencil Design, Gottschaldt Figures, and the Hanfmann-Kasanin Concept Formation Test.

The third factor, J, was called a *"judgment"* factor. It involves discrimination, ability to weigh and evaluate, to exercise control in one's thinking, and to function wisely as in a mature person. It had weight in the Reasoning test of Thurstone, the Binet Vocabulary test, the Practical Judgment test, the Word Sorting test, the Opposites test, the Stencil Design test, and the Buttons test.

A factor, P, called *"productiveness"* was primarily concerned with speed of output, although ideational richness played some role. It had weights in four of the five Thurstone tests, as well as the Gottschaldt Figures test, Beck's *Z* score, and Block Designs.

A factor, C, called *"concept"* ability was weakly identified. It apparently involves the ability to recognize essential similarities, to abstract and generalize, to think inductively. It appears to be somewhat restricted to tasks in which a considerable freedom of choice is allowed. It found expression in the battery in a test of Similarities, Comprehension, Word Sorting, Gottschaldt Figures, Holsopple Concept Formation Test, and the Buttons test.

Factor Rp was described as *"responsiveness."* It involved lack of criticism, some suggestion of willingness to please, and was felt to be partly temperamental and partly cognitive. Comprehension, Binet Vocabulary, Naming, Picture Titles, and the Holsopple test had appreciable weight in this parameter.

Factor B was the unidentified factor.

Factor Fx was labeled mental *"flexibility,"* or the ability to change one's mental set, or plasticity, or ability to shift. It was identified in Thurstone's Space test, Gottschaldt Figures, Similarities, Binet Vocabulary, Picture Anomalies, Picture Titles, and Buttons.

An attempt to validate the factors found in this study by an empirical technique based upon explanation of anomalous test scores was slightly suggestive of some clinical merit for our factors. However, the results were sufficiently ambiguous in this respect to state that the case has not been proven and further investigation is necessary along these lines.

A continued exploration by means of factor analysis in the general area of reasoning was recommended in order to clarify certain of the factors which were weakly expressed in the test battery used. This applies particularly to our *"concept formation"* factor and to the *"responsiveness"* factor. Further investigation is also needed to clarify the relative roles played by speed and/or apperceptive richness in our *"productiveness"* factor. Our *"judgment"* factor needs to be somewhat more clearly defined.

Recommendations were made that in future studies specific test batteries be designed for a particular area, as in clinical practice, because it is felt that other factors found using different approaches would have limited usefulness or application for such a specific situation. The suggestion was made that test constructors need to pay greater attention to the role of test instructions because of their probable effects upon factor composition.

It is somewhat difficult to see how some of the factors found in this analysis can be further "purified." For example, our judgment factor appears to enter into a very wide range of behavior. An adequate description of test performance would not only include the judgment aspects of the test, but would also have to include the types of material used in the test itself. The elimination of nonrelevant factors would seem to be very difficult since some of our factors are ones which are apparently operative in a wide range of test performance. Perhaps a more pragmatic point of view would be to seek tests that have clinical usefulness or readily understandable psychological meaning.

As psychologists, we must be aware of the limitations of factor analysis as a procedure, as well as aware of its possibilities. The results of this study would suggest that there are many ways of classifying intellectual performance. Just as one can describe an elephant in terms of its trunk or legs or ears or tail, so we can describe only certain aspects of mental life. The only question of concern to psychologists should be, "What purpose does my description serve?" Test construction and factor analyses should be oriented to this end.

It is suggested that the parameters employed in this study be given empirical tests in the clinic to test their practical value.

BIBLIOGRAPHY

1. Army Air Forces Aviation Psychology Program Research Reports, Printed Classification Tests, Report No. 5. U. S. Government Printing Office, 1947. Edited by J. P. GUILFORD with the assistance of J. I. LACEY. Pp. 798-821, December 1942, July 1943, November 1943, September, 1944, Classification Battery; Pp. 142-155, Judgment Test; Pp. 180-190, Foresight and Planning Tests I; Pp. 180-190, Foresight and Planning Tests II; Pp. 215-225, Integration Tests; Pp. 333-339, Mechanical Tests; Pp. 261-268, Memory Tests I; Pp. 261-268, Memory Tests II; Pp. 408-418, Perceptual Battery I; Pp. 408-418, Perceptual Battery II; Pp. 113-122, Reasoning Tests.
2. ARTHUR, G. A non-verbal test of logical thinking. *J. consult. Psychol.* 1944, **8**, 33-34.
3. ARTHUR, G. *Stencil design test, I, II.* New York: Psychological Corporation, 1944.
4. BECK, S. J. *Rorschach's test. I. Basic processes.* New York: Grune & Stratton, 1944.
5. BLAKEY, R. I. A factor analysis of a non-verbal reasoning test. *Educ. psychol. Measmt,* 1941, **1**, 187-198.
6. BURT, CYRIL. *The factors of the mind.* New York: Macmillan, 1941.
7. CARDALL, ALFRED J. *Test of practical judgment.* Chicago: Science Research Associates, 1942.
8. CORTER, H. M. *A factor analysis of some individually administered reasoning tests.* Unpublished Ph.D. dissertation. Pennsylvania State College, 1949.
9. DAVIDSON, WILLIAM M., & CARROLL, JOHN B. Speed and level components in time limit scores: A factor analysis. *Educ. psychol. Measmt,* 1945, **5**, 411-427.
10. FRENCH, JOHN W. *The factorial composition of aptitude and achievement tests* (Prepublication Draft). Lithographed. Educational Testing Service, January 1949.
11. GOODMAN, C. H. *A study of the Thurstone Primary Mental Abilities Tests.* Unpublished M.A. thesis. Pennsylvania State College, 1939.
12. GOODMAN, C. H. Factorial analysis of Thurstone's seven primary abilities. *Psychometrika,* 1943, **8**, 121-129.
13. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
14. HALSTEAD, WARD C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
15. HALSTEAD, WARD C. A power factor (P) in general intelligence: The effect of brain injuries. *J. Psychol.,* 1945, **20**, 57-64.
16. HANFMANN, E., & KASANIN, J. *Concept formation test.* Chicago: C. H. Stoelting Co., 1940.
17. HEBB, D. O. The McGill Picture Anomaly Series: Data on 100 unsophisticated adults. *Bull. Canad. Psychol. Assn.,* 1941, **1**, 47-49.
18. HEBB, D. O. The McGill Picture Anomaly Series, M and N. Orange Park, Fla.: Yerkes Laboratories, 1942.
19. HEBB, D. O., & MORTON, N. W. The McGill Adult Comprehension examination: Verbal Situation and Picture Anomaly Series. *J. educ. Psychol.,* 1943, **34**, 16-25.
20. HESTON, J. C. A factor analysis of some clinical performance tests. *J. appl. Psychol.,* 1943, **27**, 135-149.
21. HOLSOPPLE, J. Q. *Concept formation test* (unpublished).
22. MEILI, R. L'analyse de l'intelligence (The analysis of intelligence). *Arch. Psychol., Genève,* 1946, **31**, 1-64.
23. MOFFIE, D. J. A non-verbal approach to the Thurstone Primary Mental Abilities. *J. gen. Psychol.,* 1942, **27**, 35-61.
24. MYERS, C. S. A new analysis of intelligence: a critical notice. *Occup. Psychol.,* 1947, **21**, 17-23.
25. PEARSON, K. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. Roy. Soc.* **CC.,** 1902, A. 1-66.
26. RIMOLDI, H. J. A. Study of some factors related to intelligence. *Psychometrika,* 1948, **13**, 27-46.
27. TERMAN, L. M., & MERRILL, M. A. *Measuring Intelligence.* Boston: Houghton Mifflin, 1937.
28. THOMSON, GODFREY N. *A factorial analysis of human ability* (2nd rev. ed.). London: Univer. of London Press, 1946.
29. THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. of Chicago Press, 1944.
30. THURSTONE, L. L. Experimental study of simple structure. *Psychometrika,* 1940, **5**, 153-168.
31. THURSTONE, L. L. *Gottschaldt figures.* Chicago: Science Research Associates, 1947.
32. THURSTONE, L. L. *Multiple-factor analysis; a development and expansion of the vectors of the mind.* Chicago: Univer. of Chicago Press, 1947.
33. THURSTONE, L. L. Primary mental abilities. *Psychometric Monog.,* 1938, No. 1, Chicago: Univer. of Chicago Press, 1938.
34. THURSTONE, L. L. The effects of selection in factor analysis. *Psychometrika,* 1945, **10**, 165-198.

35. THURSTONE, L. L. Theories of intelligence. *Sci. Mon.*, New York, 1946, **62**, 101-112.
36. THURSTONE, L. L. *The vectors of the mind.* Chicago: Univer. of Chicago Press, 1935.
37. THURSTONE, L. L., & THURSTONE, G. T. *Factorial studies of intelligence.* Chicago: Univer. of Chicago Press, 1941.
38. THURSTONE, L. L. *SRA Primary Mental Abilities, Intermediate—Form AH.* Chicago: Science Research Associates, 1947.
39. WECHSLER, D. *Measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.
40. WECHSLER, D. *Bellevue intelligence scale Form II. Manual.* New York: Psychological Corporation, 1946.
41. WELLS, F. L. & RUESCH, J. *Mental examiner's handbook.* New York: Psychological Corporation, 1942.
42. WOLFLE, DAEL. Factor analysis to 1940. *Psychometric Monogr.*, No. 3, Univer. of Chicago Press, Chicago, 1940.